

# Expressed sequence tags analysis revealing the taxonomic position and fatty acid biosynthesis in an oleaginous green microalga, *Myrmecia incisa* Reisigl (Trebouxiophyceae, Chlorophyta)

OUYANG LongLing<sup>1</sup>, DU DaoHai<sup>1</sup>, YU ShuiYan<sup>1</sup>, LI ChunYang<sup>1</sup>, ZHANG ChengWu<sup>2</sup>, GAO HongJian<sup>3</sup> & ZHOU ZhiGang<sup>1\*</sup>

<sup>1</sup> Key Laboratory of Genetic Resources and Utilization in Aquaculture, Shanghai Ocean University, Shanghai 201306, China;

<sup>2</sup> College of Life Sciences, Ji'nan University, Guangzhou 210009, China;

<sup>3</sup> Shanghai Medical College of Fudan University, Shanghai 200030, China

Received October 19, 2011; accepted March 19, 2012; published online May 19, 2012

cDNA library of *Myrmecia incisa* H4301 was constructed using  $\lambda$  phage vectors. The library consisted of  $1.5 \times 10^6$  clones with a recombination rate of 90%, and 1942 clones were randomly sequenced. All 1854 readable expressed sequence tags (ESTs) were clustered into 596 non-redundant sequences (NRSs), among which 126 NRSs were from 1384 ESTs, showing a high redundancy. Among the 596 NRSs, 30 were ribosomal RNA, and 152 significantly matched with those available in NCBI database and JGI Genome Portal, the latter were divided into nine subcategories. Overall, 59 NRSs were involved in photosynthesis, the respiratory electron transport chain, ATP synthesis, oxidation reduction, fatty acid biosynthesis, glucose metabolism, protein metabolism, and small molecular metabolism, suggesting that these genes were abundantly transcribed during energy and substance metabolism. Acyl-carrier protein, ferredoxin and fatty acid elongase genes obtained from this cDNA library enabled presumption of a possible biosynthesis pathway of ArA in *M. incisa*. Codon usage analysis of 142 NRSs with 17798 codons in the predicted coding regions showed that the average G+C content level of the total codons was 55.39%, and that of the third position in base trimers was 66.42%, indicating a strong bias toward cytosine and/or guanosine in this algal genome. Among all synonymous codons, NAG was most favored, while NUA was most avoided. Phylogenetic trees inferred from ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit genes and the extra partial sequences of 18S rRNA obtained from this library demonstrated that *M. incisa* belonged to Trebouxiophyceae and was significantly clustered with *M. incisa* SAG 2007, *Lobosphaera tirolensis*, *M. bisecta*, and *Parietochloris incisa*, but was clearly distant from *P. pseudoalveolaris* and *P. alveolar*. Transmission electron microscopy revealed pyrenoids traversed by many parallel thylakoids membranes, while starch grains were only clearly observed when cells were grown under nitrogen stress. Based on combined investigation of the phylogeny and morphological characteristics, it is proposed that *M. incisa* be kept in the genus *Myrmecia* in which there might be two parallel groups, one living freely and another symbiotic species.

## *Myrmecia incisa*, expressed sequence tags (ESTs), codon usage, Trebouxiophyceae, phylogenetic analysis

**Citation:** Ouyang L L, Du D H, Yu S Y, et al. Expressed sequence tags analysis revealing the taxonomic position and fatty acid biosynthesis in an oleaginous green microalga, *Myrmecia incisa* Reisigl (Trebouxiophyceae, Chlorophyta). Chin Sci Bull, 2012, 57: 3342–3352, doi: 10.1007/s11434-012-5159-2

*Myrmecia incisa* Reisigl, a coccoid green microalga, is one species of Trebouxiophyceae, Chlorophyta [1]. The taxonomic position of this organism has been the subject of debate since it was first characterized. In the Culture Collec-

tion Algae of Charles University of Prague (CAUP), strain H4301 is named *Myrmecia incisa* Reisigl based on the description by Reisigl [1]. However, in Romania the same organism is named *Lobococcus incisus* (Reisigl) Reisigl, which is a synonym of *M. incisa* [2]. From observation of pyrenoids and zoospores with counterclockwise basal body

\*Corresponding author (email: zgzhou01@163.com)

orientation, Watanabe et al. [3] proposed that a strain of *M. incisa* isolated from the soil of Mt. Tateyama in Japan be reclassified into a different genus, *Parietochloris incisa* (Reisigl) Watanabe comb. nov (Trebouxiophyceae, Chlorophyta). Based on the phylogeny of the rRNA sequence and the presence of a UV-absorbing mycosporine-like amino acid, Karsten et al. [4] suggested that this species be designated as *Lobosphaera incisa* (Reisigl) U. Karsten in Trebouxiophyceae.

Recently, *M. incisa* was found to be an oleaginous alga that can accumulate an unprecedentedly high level of arachidonic acid (ArA, 20:4 $\omega$ 6)-rich triacylglycerols (TAG) in cytoplasmic lipid bodies [5–11]. Under nitrogen starvation culture conditions with a light intensity of 120  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$  for 27 d, the content of total fatty acids in *M. incisa* increased by 95% when compared with that cultured in complete BG-11 medium. ArA accounts for 7% of the dry weight biomass [10]. Among the total fatty acids in polar lipids extracted from this alga, ArA accounts for 22%, while in neutral lipids, this value increases to 50%. In addition, 76% of the intercellular ArA accumulated in the form of neutral lipids, which increased one fold in this alga when grown under nitrogen starvation conditions (unpublished data). Therefore, this species could be regarded as a potential candidate for commercial production of ArA, which is a pharmaceutically valuable long-chain polyunsaturated fatty acid (PUFA), by use of massive culture with photobioreactors [12–14]. This species also has the potential for use in studies of the metabolic pathway of ArA.

Labeling with [1- $^{14}\text{C}$ ] oleic acid has been used to elucidate the biosynthesis of ArA in *M. incisa* [15]. This technique has shown that phosphatidylcholine (PC) and diacylglycerol-trimethylhomoserine (DGTS) are the major lipids involved as acyl carriers for the  $\Delta 12$  and  $\Delta 6$  desaturation of oleic acid, which lead sequentially to linoleic acid and  $\gamma$ -linolenic acid, respectively. The latter is released from their lipid carriers and elongated to 20:3 $n$ -6, which is primarily reincorporated into phosphatidylethanolamine (PE) and PC and finally desaturated to ArA. However, there is little other information available about the metabolic mechanism of fatty acids in *M. incisa*, especially at the molecular level [11,16,17].

Expressed sequence tags (ESTs), which are generated by sequencing randomly selected clones from a cDNA library, have been shown to be very useful for functional comparisons of expressed gene populations and a powerful tool for gene discovery in the model organism *Chlamydomonas reinhardtii* [18–22]. EST analyses are especially useful when knowledge about the genome under investigation is limited, such as in the case of the green microalgae, *Scherffelia dubia* [23], *Mesostigma viride* [24], *Ostreococcus lucimarinus* [25], *Nannochloropsis oculata* [26] and *Dunaliella salina* [27]. Catalogs of ESTs are also valuable resources for the development of molecular tools and analysis methods. Sequences can be used for microarray analysis,

molecular systematics, vector construction and primer design.

Hence, to obtain more information about *M. incisa*, especially regarding the mechanism of PUFA biosynthesis and accumulation under stressed conditions, a cDNA library was constructed from a mixture sample of this alga grown under normal and stressed (nitrogen starvation) conditions. Additionally, the systematic position of this alga is discussed based on the phylogenies inferred from the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and partial 18S rRNA gene sequences obtained from the generated cDNA library.

## 1 Materials and methods

### 1.1 Algal species and cultivation conditions

The microalga, *Myrmecia incisa* H4301, was commercially provided by CAUP. This alga was cultivated in BG-11 medium [28] in 800-mL glass flasks, which were placed in a temperature-regulated photoincubator at 25°C and illuminated from the side with a light: dark regime of 12 h:12 h by Phillips (Amsterdam, Holland) cool-white fluorescent tubes (40 W) at a light irradiance of 115  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$  [6]. During culture, the flasks were shaken several times a day by hand at regular intervals. Algal cells were harvested in the late logarithmic growth phase by centrifugation at 2300 $\times g$ , after which they were washed three times with nitrogen-free fresh BG-11 medium in which ferric ammonium citrate was substituted with ferric citrate. The collected algal cells were divided into two equal portions for subsequent treatment in an attempt to maximize the variety of expressed genes, especially those related to fatty acid synthesis and accumulation of neutral lipids. One portion was re-suspended in normal medium, while the other was cultivated in nitrogen-free medium. Both cultures were then incubated under the conditions described above for another 7 d before harvest. The samples were then washed with fresh media with or without nitrogen, mixed together and stored in liquid nitrogen for total RNA extraction.

### 1.2 RNA extraction and cDNA library construction

Total RNA was extracted using TRIzol reagent (Invitrogen, USA). A non-normalized cDNA library was constructed using the SMART<sup>TM</sup> cDNA Library Construction Kit (Clontech, USA) according to the manufacturer's instructions. Complementary DNA synthesis was conducted using long-distance PCR (LD PCR) [29] to generate full-length cDNA. After digestion with *Sfi* I and size fractionation using CHROMA SPIN<sup>TM</sup> Columns, cDNAs (>0.4 kb) were ligated to the *Sfi* I-digested  $\lambda$ TriplEx2<sup>TM</sup> vectors. Subsequently,  $\lambda$  phages were packaged and transformed into XL1-Blue *Escherichia coli* competent cells. Prior to the amplification of the original library on plates at 37°C, the

primary library titer, recombination ratio and transformation ratio were calculated. Samples were stored at  $-80^{\circ}\text{C}$  in LB plates containing 25% glycerol until use.

### 1.3 Sequencing and data analysis

EST sequencing was performed from the 5'-terminus by Shanghai Biochip Co., Ltd. (Shanghai, China) using an ABI3700 sequencer. Data from the sequencer were processed manually. Vector sequences were removed, and ambiguities were resolved with reference to original trace data using BioEdit (Ibis Biosciences, Carlsbad, CA, USA) and the VecScreen program (<http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html>). High-quality ESTs with a minimum of 100 bp of continuous sequences and at least 98% accuracy were assembled into non-redundant sequences (NRSs) (contigs and singletons) using the DNAMAN Program (Lynnon Biosoft, Quebec, Canada), with a threshold of minimum overlap of 100 bp and identity of 95%. The percentage redundancy was calculated as follows:  $100 \times [\text{number of ESTs} - \text{non-redundant sequences (NRSs)}] / \text{total number of ESTs}$  [30]. Each NRS was compared as 3-frame translations to the NCBI non-redundant protein database using blastx (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Similarities to known proteins were regarded as significant if the probability values (*E*-values) were less than 0.0001, the identity was greater than 40%, and the region of similarity was more than 50 amino acids in length as suggested by Scala et al. [31]. Otherwise, the tblastx, blastn and blastx alignment programs were used to search the JGI Genome Portal (<http://genome.jgi-psf.org/>) for the remaining unknown NRSs after searching the NCBI database. The recovered species included *Arabidopsis lyrata*, *Chlamydomonas reinhardtii*, *Chlorella* sp. NC64A, *Coccomyxa* sp. C-169, *Micromonas* sp. RCC299, *Ostreococcus lucimarinus*, *Ostreococcus* sp. RCC809, *Ostreococcus tauri* and *Volvox carteri* f. *nagariensis* when the database was searched using default parameters for all models (transcripts). Functional classification was manipulated by GO (<http://www.geneontology.org/>), and codon usage and G/C-content analysis was accomplished from 142 NRSs using the ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf.html>) and Codon W program [32]. Sequences detailed in the present study have been deposited in the NCBI database.

### 1.4 Construction of phylogenetic trees

Although the rRNA sequence was not expected to be screened in a cDNA library, it could provide extra information regarding the taxonomic position of *M. incisa* analysis. Two independent types of data analyses were used to assess the evolutionary relationships resolved in the rRNA and ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (*rbcS*) phylogenies with the neighbor-joining (NJ) and maximum parsimony (MP) methods in MEGA 4 [33].

The 18S rRNA sequence and *rbcS* amino acid sequence of *M. incisa* were obtained from the cDNA library, while sequences used for comparison were retained from GenBank. All accession numbers are presented in the phylogenetic trees. Homologous sequences were aligned using Clustal X [34] with the default parameter settings. During NJ analysis, the Tamura and Nei model was selected. Support for internal branches in the MP and NJ trees was estimated using the bootstrap method [35]. Bootstrap resampling was performed with 1000 replications for both the NJ and MP trees.

### 1.5 Ultrastructure observation of *M. incisa* by transmission electron microscopy (TEM)

Actively growing *M. incisa* cells cultured in complete BG-11 medium and nitrogen-free BG-11 medium were collected during the exponential phase. Samples for TEM were prepared using the modified protocol of Campbell and Plank [36]. Briefly, the samples were subject to prefixation in 2.5% glutaraldehyde in  $0.1 \text{ mol L}^{-1}$  phosphate buffer at pH 7.4 for more than 2 h, after which they were washed with  $0.1 \text{ mol L}^{-1}$  phosphate buffer three times. Next, the samples were subject to post-fixation (3 h) with 1% osmium tetroxide in  $0.1 \text{ mol L}^{-1}$  phosphate buffer at pH 7.4, and then washed as described above. The samples were subsequently dehydrated in a graded series of ethanol and two times in 100% acetone. Next, the samples were placed in acetone/Spurr's resin (2/1, v/v) for 3–4 h, and then in acetone/Spurr's resin (1/2, v/v) overnight, and finally incubated in 100% Spurr's resin for 2–3 h at  $37^{\circ}\text{C}$ . Finally, the embedded samples were polymerized in a temperature gradient and thin-sectioned using an ultracut microtome (LKB Ultratome 4802; Leica, Germany). The sections were then stained with 3% uranyl acetate-lead citrate, and observed and photographed using a JEOL JEM-1230 Transmission Electron Microscope at 80 kV.

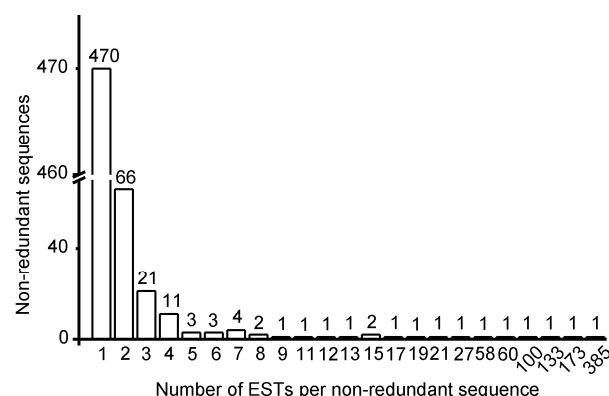
## 2 Results

### 2.1 EST cluster analysis of cDNA library from *M. incisa*

Single-pass nucleotide sequences were obtained from a total of 1942 clones selected at random from the cDNA library of *M. incisa*, which consisted of  $1.5 \times 10^6$  clones, approximately 90% of which were recombinant. Overall, 4.53% (88 ESTs) of these sequences were found to contain no insert or be too short for analysis ( $<100 \text{ bp}$ ), and thus were not considered. The remaining 1854 readable sequences comprised 596 non-redundant sequences (NRSs), which included 470 tentative unique singletons (TUSs) and 126 tentative unique contigs (TUCs). The latter was assembled from the remaining 1384 ESTs (Table 1), showing a relatively high redundancy. This might have been because the constructed cDNA library is a non-normalized cDNA library, which is much

more redundant but could provide raw information regarding the structure of gene expression levels [37–39]. The lengths of the TUSs and TUCs ranged from 106 to 1026 bp with an average of 483 bp and from 103 to 1278 bp with an average of 571 bp, respectively, indicating that the EST sequences were credible and suitable for further analysis.

The 126 TUCs contained a minimum of 2 ESTs and a maximum of 385 ESTs (Figure 1). Approximately 52.4% of the TUCs had 2 ESTs, 25.4% had 3–4 ESTs, and 22.2% had more than 5 ESTs. The TUCs made from the largest number of ESTs from *M. incisa* were the large (385 ESTs) or small (173 ESTs) subunits of chloroplast ribosomal RNA genes (Table 2), followed by the large subunits of mitochondrion (133 ESTs) and chloroplast (60 ESTs) ribosomal RNA genes. The rest of the ten most abundant TUCs, which



**Figure 1** Distribution of the number of expressed sequence tags (ESTs) per assembled non-redundant sequence in *Myrmecia incisa*. In total, 1854 EST sequences were analyzed.

**Table 1** Summary of the sequencing and clustering of the *Myrmecia incisa* cDNA library

Non-normalized cDNA library	Number	Percent (%)
Total clones	1942	
Unreadable sequences	88	4.53
Total ESTs analyzed	1854 (788 ESTs are rRNA)	95.47
Non-redundant sequences (NRSs)	596 (30 NRSs are rRNA)	
Tentative unique contigs (TUCs)	126 (5 are rRNAs) containing 1384 ESTs (763 are rRNAs)	21.14
Tentative unique singletons (TUSs)	470 (25 are rRNAs)	78.86
Significantly matched NRSs <sup>a)</sup>	152	26.86
No significantly matched NRSs <sup>a)</sup>	414	73.14

a) Excluding 30 NRSs that were rRNA sequences

**Table 2** Putative functions of the largest contigs including more than 10 cDNAs in the cDNA library of *Myrmecia incisa*

Sequence name	Number of cDNAs	Length (bp)	Putative identification (organism)	Identity (%)	E-value	Score
Contig001	385	854	Chloroplast 23S rRNA/rrn1	87	0.0	1002
Contig002	173	451	Chloroplast 16S rRNA	98	0.0	756
Contig003	133	1109	Mitochondrion 23S rRNA/rrn1	73	$4 \times 10^{-127}$	464
Contig004	100	644	No significant homology	—	—	—
Contig006	60	515	Chloroplast 23S rRNA/rrn1	86	$1 \times 10^{-90}$	342
Contig005	58	899	No significant homology	—	—	—
Contig007	27	1100	RuBisCO small subunit 1 of ( <i>Acetabularia acetabulum</i> )	65	$5 \times 10^{-53}$	212
Contig126	21	662	No significant homology	—	—	—
Contig008	19	953	No significant homology	—	—	—
Contig122	17	884	No significant homology	—	—	—
Contig009	15	537	Putative cruciform DNA binding protein	54	$5 \times 10^{-23}$	110
Contig010	15	739	No significant homology	—	—	—
Contig011	13	241	No significant homology	—	—	—
Contig012	12	588	Mitochondrion 23S rRNA	77	$2 \times 10^{-89}$	338
Contig125	11	610	No significant homology	—	—	—
Contig013	9	640	Plant Late Embryo Abundant (LEA) related family member	—	—	—
Contig014	8	546	Cytochrome c oxidase subunit 2 ( <i>Chlorokybus atmophyticus</i> )	75	$4 \times 10^{-52}$	207
Contig015	8	1209	No significant homology	—	—	—
Contig016	7	1278	NADH dehydrogenase (ubiquinone), subunit 4 ( <i>Prototheca wickerhamii</i> )	81	$6 \times 10^{-175}$	617
Contig017	7	1240	No significant homology	—	—	—
Contig018	7	104	No significant homology	—	—	—
Contig022	7	642	Late embryogenesis abundant protein-like	—	—	—
Contig019	6	428	No significant homology	—	—	—
Contig020	6	544	No significant homology	—	—	—
Contig023	6	490	Photosystem II complex subunit J ( <i>Pseudodoclonium akinetum</i> )	89	$1 \times 10^{-6}$	55.5

consisted of 100, 58, 21, 19 and 17 ESTs, respectively, were all genes of unknown function, except for one (27 ESTs) encoding a ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (*rbcS* gene). The contamination of these abundant ribosomal RNAs may explain why the redundancy of the cDNA library was as high as 67.85%. Most of these ESTs were deposited in the GenBank public database under accession Nos. FE673534–FE673550, FF848255–FF848339 and FG065108–FG065165. Because some of these ESTs might represent non-overlapping sequences of the same genes, the total likely represented an overestimate of the actual number of unique ESTs.

## 2.2 Functional annotation and categorization

Excluding 30 NRSs that were similar to rRNA genes while searching in the nucleotide sequence, the rest were classified into two categories based on their putative functions: significant matches (26.86%) to sequences in the public database, and no significant matches (73.14%) (Figure 2). The 152 NRSs with significant matches were divided into nine subcategories: signal transduction (2.63%), stress response (1.31%), substance metabolism (11.18%), energy metabolism (27.63%), genetic translation factors (28.95%) including the large and small subunits of the ribosomal protein, genetic transcription factors (4.61%), substrate transmembrane transport (4.61%), other functions (3.29%) and unknown proteins containing hypothetical proteins, predicted proteins and unnamed proteins (15.79%).

As shown in Table S1, the NRSs in the subcategory of substance metabolism were mainly involved in glucose (2 NRSs), small molecule (7 NRSs), protein (4 NRSs) and fatty acid biosynthesis (4 NRSs), while those in the subcategory of energy metabolism were primarily involved in photosynthesis (15 NRSs), oxidation reduction (14 NRSs), ATP synthesis (9 NRSs) and the respiratory electron transport chain (4 NRSs). Highly significant matches most frequently obtained sequences from algae (79 NRSs) and

higher plants (47 NRSs) (Figure 2). Ten NRSs showed the highest similarity to bacteria, eight to fungi, and the rest had significant matches to animals and other sequences, such as synthetic constructs (Figure 2).

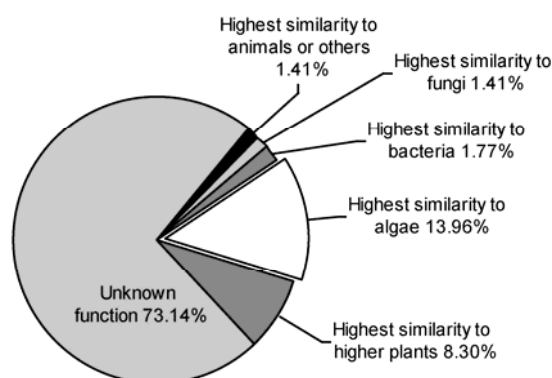
Four NRSs directly involved in fatty acid metabolism and accumulation were identified (Table S1). One TUS 535 bp (SZ\_06\_07) and another 431 bp in length (GenBank accession No. FF848273) had significant similarity with the acyl-carrier protein gene (identity, 76%; *E*-value,  $3 \times 10^{-24}$ ) and phosphatidylglycerolphosphate synthase gene (identity, 81%; *E*-value,  $2 \times 10^{-18}$ ) of *Chlamydomonas reinhardtii*, respectively, indicating that these genes involved in the phospholipid synthesis pathway of *M. incisa* might be similar to *C. reinhardtii*. The other two TUS had 56% (*E*-value,  $6 \times 10^{-10}$ ) and 92% (*E*-value,  $1 \times 10^{-5}$ ) similarity to polyunsaturated fatty acid elongase from *Marchantia polymorpha* L. and carboxylesterase from *Marinobacter algicola* DG893, respectively.

One NRS (GenBank accession No. FF848334) 600 bp in length was identified as putative ferredoxin, which is considered an electron donor for double bond introduction into fatty acids by acyl-lipid desaturases in cyanobacterial cells and chloroplasts.

## 2.3 Codon usage and G/C-content analysis

A codon usage analysis of 142 NRSs (100 TUSs and 42 TUCs) with 17798 codons in the predicted coding regions indicated a strong bias (>60%) for guanosine and/or cytosine at the third position of triplet codons for Leu, Ile, Val, Tyr, His, Gln, Asp, Lys, Asn, Glu, Cys, Arg and Gly. A weaker G/C preference (>50%) was observed for several hydrophobic amino acids, including Phe, Ser, Thr and Ala, but no bias was found for Pro (Table 3). If there was a choice between C and G, cytosine was preferred for the amino acids mentioned above, except for the two hydrophobic amino acids Leu and Val. For the entire collection of NRSs, the percentage occupation of the third position of all codons was U=18.04%, A=14.13%, C=35.13% and G=32.70%. Additionally, the results suggest that this alga prefers to use UAG as a stop codon, as indicated by usage ratios of UAA, UAG and UGA of 26.51%, 39.76% and 33.73%, respectively. Synonymous codon bias showed that NAG (N representing A, U, G and C) was most favored, while NUA was most avoided.

The average level of total G+C content of these 142 NRSs was 55.39%, of which 28.87% (at a frequency of 41) were higher than 60%, while the average level of G+C content at the third position was 66.42%, of which 82.39% (at a frequency of 117) were higher than 60% (Figure 3). Based on the G/C content, amino acids were divided into a high G/C content class (>66%), composed of Pro, Ala, Gly and Trp, a middle G/C content class (>50%) containing Val, Thr, His, Glu, Asp, Gln, Cys and Ser, and a low G/C content class consisting of Phe, Tyr, Asn, Lys, Ile and Met. Arg and

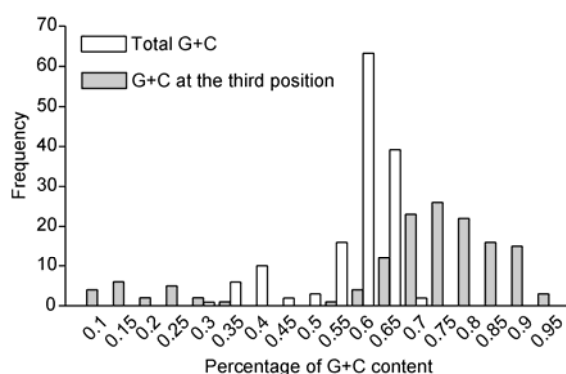


**Figure 2** Distribution of matches in the non-redundant protein databases. These sequences from *Myrmecia incisa* shared similarity with those from bacteria, fungi, algae, higher plants, or animals.

**Table 3** Codon usage in the *Myrmecia incisa* ESTs<sup>a)</sup>

Codon	Occurrence frequency	Percent (%)	Codon	Occurrence frequency	Percent (%)	Codon	Occurrence frequency	Percent (%)	Codon	Occurrence frequency	Percent (%)
UUU F	318	41.95	UCU S	182	15.84	UAU Y	184	37.02	UGU C	78	26.09
UUC F	440	<u>58.05</u>	UCC S	224	19.50	UAC Y	313	<u>62.98</u>	UGC C	221	<u>73.91</u>
UUA L	167	10.12	UCA S	180	15.67	UAA*	22	26.51	UGA*	28	33.73
UUG L	228	13.82	UCG S	128	11.14	UAG*	33	39.76	UGG W	219	100
CUU L	174	10.55	CCU P	206	25.03	CAU H	141	37.70	CGU R	170	16.07
CUC L	239	14.48	CCC P	271	<u>32.93</u>	CAC H	233	<u>62.30</u>	CGC R	382	<u>36.11</u>
CUA L	91	5.52	CCA P	215	26.12	CAA Q	175	24.75	CGA R	91	8.60
CUG L	751	<u>45.52</u>	CCG P	131	15.92	CAG Q	532	<u>75.25</u>	CGG R	204	19.28
AUU I	217	25.09	ACU U	200	19.76	AAU N	118	19.73	AGU S	100	8.70
AUC I	537	<u>62.08</u>	ACC U	357	<u>35.28</u>	AAC N	480	<u>80.27</u>	AGC S	335	<u>29.16</u>
AUA I	111	12.83	ACA U	257	25.40	AAA K	228	18.51	AGA R	87	8.22
AUG M	470	100	ACG U	198	19.57	AAG K	1004	<u>81.49</u>	AGG R	124	11.72
GUU V	201	15.29	GCU A	391	24.79	GAU D	261	29.23	GGU G	270	20.22
GUC V	375	28.52	GCC A	511	<u>32.40</u>	GAC D	632	<u>70.77</u>	GGC G	703	<u>52.66</u>
GUA V	122	9.28	GCA A	371	23.53	GAA E	196	22.17	GGA G	174	13.03
GUG V	617	<u>46.92</u>	GCG A	304	19.28	GAG E	688	<u>77.83</u>	GGG G	188	14.08

a) Analysis is of 142 proteins (17798 codons). The third position nucleotide usage was U = 18.04%, A = 14.13%, C = 35.13%, G = 32.70%. The superscript \* indicates a stop codon and the underlined number displays the highest frequency in degenerate codons. The capital letter following the codon indicates the corresponding amino acid.

**Figure 3** Distribution of 142 NRSs that had significant matches in the database and contained an ORF region according to G+C content. Contaminating ribosomal RNA sequences were not included in this analysis.

Leu were not assigned to any of these classes because Arg consisted of both a high G/C content class and a middle G/C content class, while Leu consisted of both a middle G/C content and low G/C content class. Amino acid cost analysis showed that the top five amino acid components in the aforementioned putative proteins were Leu, Ala, Val, Gly and Lys, which was consistent with the total G/C and the third position G/C contents (data not shown).

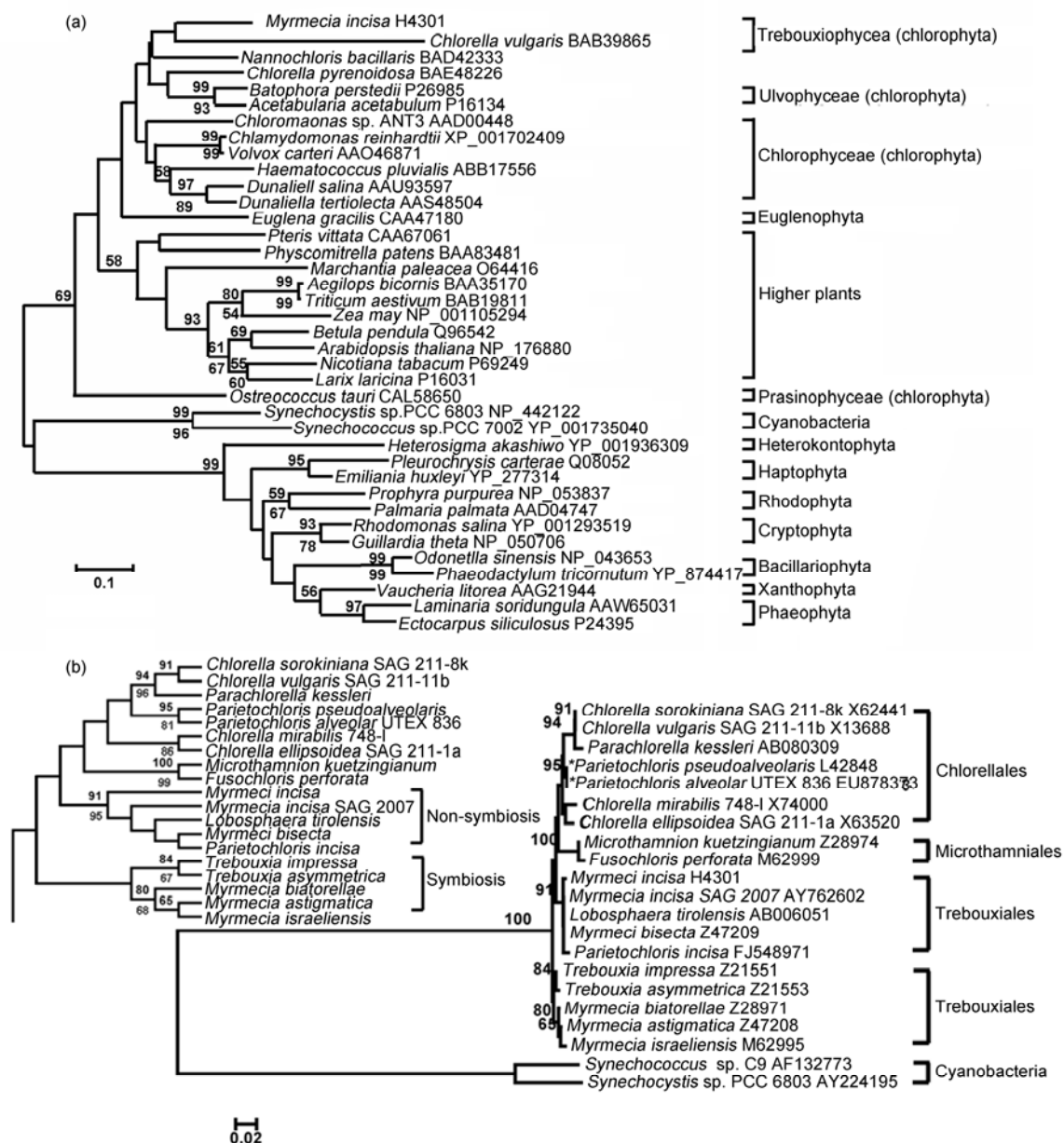
## 2.4 Phylogenetic analysis

A putative full-length RuBisCO small subunit gene (GenBank accession No. EU797508) with abundant transcripts (27 ESTs) was obtained from the cDNA library. The gene consisted of a 76-bp 5'-untranslated region (UTR), a 462-bp 3'-UTR and a 570-bp open reading frame (ORF) encoding 189 amino acids. The gene had 66% identity to the chloro-

plast *rbcs* from *Chlorella pyrenoidosa* with an *E*-value of  $6 \times 10^{-59}$ . Based on the deduced peptide sequences of *rbcs* of this alga and other selected species, a NJ phylogenetic tree was constructed using a set of MP tree bootstrap values to confirm the results (Figure 4(a)).

*M. incisa* was first clustered with *Chlorella* and *Nannochloris* constituting an independent lineage within the Trebouxiophyceae, which was different from Ulvophyceae and Chlorophyceae in Chlorophyta (Figure 4(a)). These three classes were then clustered with Euglenophyta and higher plants, suggesting that the *rbcs* gene in this microalga was encoded by nuclear genome as in other chlorophytes and higher plants. Four classes of Chlorophyta were resolved in this constructed phylogenetic dendrogram, but Prasinophyceae including *Ostreococcus tauri* was separated by Euglenophyta and higher plants. The clade consisting of Chlorophyta, Euglenophyta and higher plants was distinct from the other clade with a bootstrap value of 69 (Figure 4(a)). The latter included cyanobacteria, Heterokontophyta, Haptophyta, Rhodophyta, Cryptophyta, Xanthophyta, Bacillariophyta and Phaeophyta, in which the *rbcs* gene was encoded by the plastid genome.

A partial nuclear-encoded small-subunit ribosomal RNA (18S rRNA) gene (GenBank accession No. FJ384550) 472 bp in length was annotated to the 18S ribosomal RNA gene from *M. incisa* strain SAG 2007 with 99% identity at an *E*-value of zero. The 18S rRNA sequence comparisons demonstrated that *M. incisa* H4301 was highly related to *M. incisa* SAG 2007, *L. tirolensis*, *M. bisecta* and *P. incisa*, which was supported by NJ and MP bootstrap values of 91 and 95, respectively (Figure 4(b)). In the genus of *Myrmecia*, the other species characterized by symbiosis with lichens were also significantly clustered together at a NJ



**Figure 4** Strict consensus tree of two phylogenies independently inferred using the Neighbor-joining and maximum parsimony methods for analysis of the *rbcS* (a) and 18S rRNA (b). Bootstrap values were computed independently for 1000 resamplings of NJ analysis (above the line) and 1000 resamplings of MP analysis (below the line). Values lower than 50 are not shown. The inset of (b) displays only the topology of the bootstrap consensus tree. Scale indicates evolutionary distance.

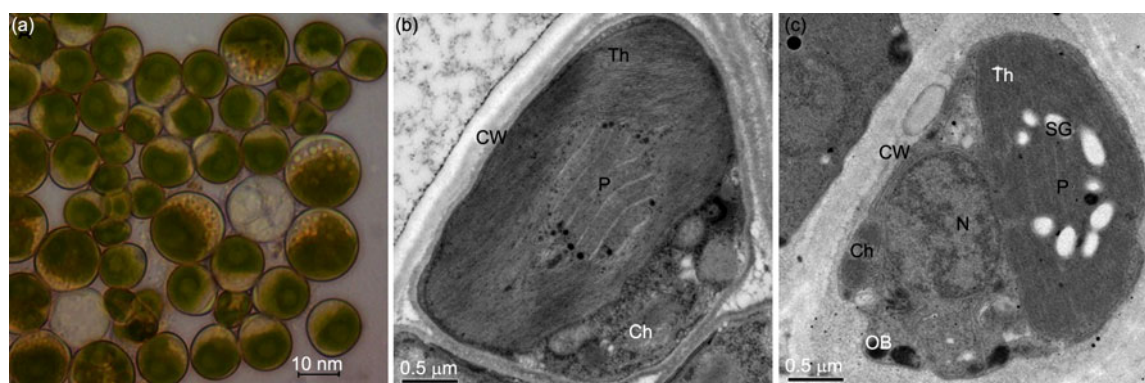
bootstrap value of 100 (Figure 4(b)). This phylogenetic tree suggested that *Myrmecia* might be a paraphyletic taxon, but this genus was distinct from the *Parietochloris* genus, which formed a cluster with *P. pseudoalveolaris* and *P. alveolaris* as a sister group with species from Chlorellales and Microthamniales.

## 2.5 Ultrastructural characteristics of *M. incisa*

Under the light microscope, cells of *M. incisa* were spherical, ovoid and 10–15  $\mu\text{m}$  in diameter (Figure 5(a)). The chloroplast was parietal and deeply incised in several parts,

showing various morphologies among individuals. The cell wall of *M. incisa* was thick, even in spherical cells, and this thickness was greater in the narrower portion of ovoid cells. Transmission electron microscopy showed pyrenoids traversed by many parallel thylakoids membranes (Figures 5(b) and (c)). Starch grains appeared around the pyrenoid when the algal cells were cultured in nitrogen-free BG-11 medium for 15 d, but were absent when cells were in the exponential phase in the complete BG-11 medium. These findings suggest that starches were usually scarce while cells were in the normal culture medium, so that pyrenoids might easily be overlooked by light microscopy analysis of *M. incisa*.





**Figure 5** *Myrmecia incisa* morphology and ultrastructure. Morphological characteristics were observed by light microscopy (a); ultrastructural characteristics of *M. incisa* cultured in complete BG-11 medium (b) and nitrogen-free BG-11 medium for 15 d (c) were observed by transmission electron microscopy. Ch: chloroplast; P: pyrenoid; Th: thylakoids; OB: oil body; N: nucleus; CW: cell wall; SG: starch grain.

### 3 Discussion

#### 3.1 Biosynthesis of ArA in *M. incisa*

Arachidonic acid is produced from stearic acid by desaturation and elongation in this alga as suggested by Bigogno et al. [15]. However, stearic acid is formed *de novo* by condensation from 2C units. Although only 152 NRSs are annotated in this cDNA library, one (SZ\_06\_07) is homologous to the acyl carrier protein gene from *C. reinhardtii* Dang. This suggests that *M. incisa* can mediate fatty acid synthesis in the activated form of acyl-acyl carrier protein in a fashion similar to other plants, including algae and most bacteria [40].

Ferredoxin, which plays an important role as a cofactor in many biological reactions, is also considered an electron donor for double bond introduction into fatty acids by acyl-lipid desaturases in cyanobacterial cells and chloroplasts [41]. It seems reasonable to suggest that *M. incisa* chloroplasts have an essential cofactor for fatty acid desaturation because one NRS (GenBank accession No. FF848334) was highly homologous with the chloroplast ferredoxin gene from *Micromonas pusilla* CCMP1545. Unfortunately, no desaturase gene was screened from this cDNA library, possibly because there were too few sequences. However, a  $\omega$ 3 fatty acid desaturase gene was cloned from this microalga using a pair of degenerated primers, and the transcription profile of this gene showed that this desaturase plays an important role in  $\alpha$ -linolenic acid (18:3 $\omega$ 3, ALA) synthesis by allocating linoleic acid (18:2 $\omega$ 6, LA) in this way under normal culture conditions [9].

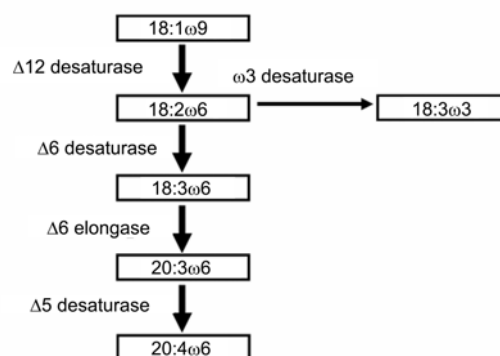
Elongation catalyzed by fatty acid elongases from 18-carbon enoyl acid is regarded as a key step in long-chain fatty acid biosynthesis [42]. Based on the clone (SZ\_07\_83) screened from this cDNA library of *M. incisa*, a fatty acid elongase gene has been cloned [11] and confirmed to elongate  $\gamma$ -linolenic acid (18:3 $\omega$ 6, GLA) and 18:4 $\omega$ 3 [43]. The transcription of this gene was found to be enhanced prior to ArA content accumulation, suggesting it may be responsible

for the ArA accumulation in this microalga cultured under nitrogen starvation conditions [11], which is consistent with the results of a study conducted by Iskandarov et al. [17]. Hence, a possible synthesis pathway of ArA in *M. incisa* was hypothesized to start from oleic acid via LA, GLA and 20:3 $\omega$ 6 to the final ArA (Figure 6).

In general, acyl-carrier protein, ferredoxin and fatty acid elongase genes enable *M. incisa* to conduct *de novo* synthesis of ArA. Therefore, it is very clear that the EST approach has enabled examination and isolation of many transcripts from *M. incisa*, which is difficult to achieve by single-gene research. In addition, these findings have provided a foundation of the biological processes involved. Moreover, the materials and information derived will be useful for preparation of gene arrays for future functional genomic research and understanding fundamental aspects of the growth and physiology of this oleaginous green microalga.

#### 3.2 Codon usage in *M. incisa*

Amino acids encoded by four or six synonymous codons in *M. incisa* favored C in the third position, excluding those encoded by NUG. This phenomenon has also been observed in *Chlamydomonas reinhardtii*, *Parachlorella kessleri*,



**Figure 6** Possible synthesis pathway of ArA in *Myrmecia incisa* deduced from the annotated NRSs.



*Ostreococcus tauri*, *Ostreococcus lucimarinus*, and *Synechococcus* sp. CC9605 (data from <http://www.kazusa.or.jp/codon/>). Shi et al. [44] indicated that human genes contained significantly biased codons, with high codon-anticodon interaction energy, in which the third position is cytosine. Therefore, the favorite C-ending codons in *M. incisa*, which has the strongest codon-anticodon binding energies, might be very helpful in guaranteeing the fidelity of genes translation. Moreover, the base trimer CUG was predominant in encoding Leu, which was one of the top five amino acids used. These findings are reasonable because Leu is a major ingredient of proteins rich in secondary structures and Leu-encoding CTG is a universally preponderant Leu codon [45].

The efficiency of translation of some codons can be modulated by the codon context, and the rules for preferable synonymous codon choice in highly expressed genes depend on the nucleotides surrounding the codon present in *Escherichia coli* [46]. Ohno [47] reported that UA/CG deficiency-UG/CU excess was a universal rule for coding sequence construction based on analysis of the coding regions of human estrogen receptor, phosphoglycerate kinase, pollen-stigma self-incompatibility protein of the mustard plant and rainbow trout H1 histone. Among all analyzed synonymous codons in *M. incisa*, NUA was most avoided as described by Ohno [47], while NAG was most favored. These findings are concordant with the results observed for *C. reinhardtii*, *P. kessleri*, *Synechococcus* sp. CC9605, and *O. lucimarinus* (data from <http://www.kazusa.or.jp/codon/>). This was probably because the systematic evolution position of algae is far from that of mammals, fish and higher plants. Different binding energies among synonymous codons might be used to suit certain situations to the translation process.

### 3.3 Taxonomy position of *M. incisa*

Seven species in the genus *Myrmecia*, *M. globosa* Printz., *M. incisa* Reisigl., *M. macronucleata* (Deason) V. andr. comb. nov., *M. biatorellae* (Tschermak-Woess et Plessl) B.-Peters., *M. bisecta* Reisigl., *M. astigmatica* Vinatzer., and *M. reticulata* Tschermak-Woess, have been described and reviewed by Andreyeva [48]. Based on the small subunit of ribosomal RNA genes, these species belong to the class Trebouxiophyceae, as indicated by Friedl [49]. In the present study, the phylogenetic tree (Figure 4(a)) inferred from the putative RuBisCO small subunit amino acid sequences provided further molecular evidence that *M. incisa* belongs to this class. Because there are no more RuBisCO small subunit amino acid sequences available from other species of *Myrmecia* in the public database (with the exception of *M. incisa*), it seems reasonable that there is no more detailed specific classification of this genus based on this constructed phylogeny (Figure 4(a)).

Since the morphological description by Reisigl in 1964

[1], the taxonomic position of *M. incisa* is still in debate [2–4]. This doubt was not well resolved by Friedl [49] using the small subunit of ribosomal RNA genes, possibly because no data was available about this species at that time. Fortunately, more data describing the 18S rRNA gene were obtained from this cDNA library, which enable further classification of *Myrmecia* [50]. The phylogenetic analysis of the partial sequence of 18S rRNA (Figure 4(b)) indicates that *Myrmecia* might be a paraphyletic taxon consisting of lichen symbionts and free-living species as described by Friedl [49]. The former includes *M. israeliensis*, *M. biatorellae*, *M. astigmatica*, *Trebouxia impressa* and *T. asymmetrica* [51], while the latter comprises *M. incisa* H4301, *M. incisa* SAG 2007, *M. bisecta*, *L. tirolensis* and *P. incisa*. The small subunit of the ribosomal RNA of *M. incisa* H4301 has one different base from that of *M. incisa* SAG 2007, as well as from *M. bisecta*, suggesting that they might be different strains of *M. incisa*. Seven different bases are found between *M. incisa* H4301 and *P. incisa*, regardless of ambiguous bases of *P. incisa* 18S rRNA, while 9 and 12 different bases are found in comparison with *P. pseudoalveolaris* and *P. alveolaris*, respectively, implying that *P. incisa* is more different from *Parietochloris* than from *Myrmecia*.

Watanabe et al. [3] proposed that *M. incisa* should be reclassified into a different genus, *Parietochloris*, because of the pyrenoids and counterclockwise basal body orientation of zoospores observed upon electron microscopy. It is necessary to point out that the counterclockwise absolute orientation of zoospore basal bodies described by Friedl [49] is a common feature of Trebouxiophyceae, so it does not seem to be useful as a criterion for identification among species in Trebouxiophyceae. Pyrenoids can be used as a criterion for the classification of algae, and *Myrmecia* were found to have pyrenoids by Metting [52]. However, it should be noted that investigation of the presence of pyrenoids by light microscopy is not highly effective because of the lack of starches [53]. Using transmission electron microscopy, pyrenoids traversed by many parallel thylakoids membranes were observed in this study, and starch grains were clearly observed only when cells were grown under nitrogen stress conditions (Figure 5), which is consistent with the results of a study conducted by Merzlyak et al. [8]. Therefore, the appearance of pyrenoids was likely overlooked by Tschermak-Woess and Plessl [54] and Reisigl [1] when they observed the cellular structure of *Myrmecia* using only light microscopes. In addition, the cell walls observed in this study were partially thickened, which differs from that of *Parietochloris* [55], suggesting that this species differs from *Parietochloris*. Taking the present morphology and phylogenetic analysis together as an integrative taxonomy as suggested by Dayrat [56], it is proposed that this species should be a member of *Myrmecia*, and *Parietochloris incisa* retain the previous name *Myrmecia incisa*.

This work was supported by the National Natural Science Foundation of China (30972243), the Creative Project of the Shanghai Municipal Education Committee (09ZZ167), the National High Technology Research and Development Program of China (2009AA064401), the Key Discipline Project of the Shanghai Municipal Education Committee (J50701), and an Excellent Thesis Grant from Shanghai Ocean University.

- Reisigl H. Zur systematik und ökologie alpiner Bodenalgae (in German). Österr Bot Z, 1964, 111: 402–499
- Cărauş I. Algae of Romania—a distributional checklist of actual algae. Doctor Dissertation. Bacău: Universitatea din Bacău, 2002
- Watanabe S, Hirabayashi S, Boussiba S, et al. *Parietochloris incisa* comb. nov. (Trebouxiophyceae, Chlorophyta). Phycol Res, 1996, 44: 107–108
- Karsten U, Friedl T, Schumann R, et al. Mycosporine-like amino acids and phylogenies in green algae: *Prasiola* and its relatives from the Trebouxiophyceae (Chlorophyta). J Phycol, 2005, 41: 557–566
- Bigogno C, Khozin-Goldberg I, Boussiba S, et al. Lipid and fatty acid composition of the green oleaginous alga *Parietochloris incisa*, the richest plant source of arachidonic acid. Phytochemistry, 2002, 60: 497–503
- Bigogno C, Khozin-Goldberg I, Cohen Z. Accumulation of arachidonic acid-rich triacylglycerols in the microalga *Parietochloris incisa* (Trebouxiophyceae, Chlorophyta). Phytochemistry, 2002, 60: 135–143
- Khozin-Goldberg I, Bigogno C, Shrestha P, et al. Nitrogen starvation induces the accumulation of arachidonic acid in the freshwater green alga *Parietochloris incisa* (Trebouxiophyceae). J Phycol, 2002, 38: 991–994
- Merzlyak M N, Chivkunova O B, Gorelova O A, et al. Effect of nitrogen starvation on optical properties, pigments, and arachidonic acid content of the unicellular green alga *Parietochloris incisa* (Trebouxiophyceae, Chlorophyta). J Phycol, 2007, 43: 833–843
- Li C Y, Du D H, Yu S Y, et al. Characterization of a  $\omega$ 3 fatty acid desaturase gene from *Myrmecia incisa* and its relative transcription during the stress course of nitrogen starvation (in Chinese). J Fish China, 2010, 34: 1343–1353
- Tong M, Yu S Y, Ouyang L L, et al. Comparison of increased arachidonic acid content in *Myrmecia incisa* cultured during the course of nitrogen or phosphorus starvation (in Chinese). J Fish China, 2011, 35: 763–773
- Yu S Y, Liu S C, Li C Y, et al. Submesoscale characteristics and transcription of a fatty acid elongase gene from a freshwater green microalgae, *Myrmecia incisa* Reisigl. Chin J Oceanol Limnol, 2011, 29: 87–95
- Zhang C W, Cohen Z, Khozin-Goldberg I, et al. Characterization of growth and arachidonic acid production of *Parietochloris incisa* comb. nov. (Trebouxiophyceae, Chlorophyta). J Appl Phycol, 2002, 14: 453–460
- Liu J G, Cohen Z, Richmond A. Fatty acids profile in a high cell density culture of arachidonic acid-rich *Parietochloris incisa* comb. nov. (Trebouxiophyceae, Chlorophyta) exposed to high PFD. Chin J Oceanol Limnol, 2002, 20: 149–156
- Wang C, Sun Y, Xing R, et al. Effect of liquid circulation velocity and cell density on the growth of *Parietochloris incisa* in flat plate photobioreactors. Biotechnol Bioproc Eng, 2005, 10: 103–108
- Bigogno C, Khozin-Goldberg I, Adlerstein D, et al. Biosynthesis of arachidonic acid in the oleaginous microalga *Parietochloris incisa* (Chlorophyceae): Radiolabeling studies. Lipids, 2002, 37: 209–216
- Iskandarov U, Khozin-Goldberg I, Cohen Z. Identification and characterization of  $\Delta$ 12,  $\Delta$ 6, and  $\Delta$ 5 desaturases from the green microalga *Parietochloris incisa*. Lipids, 2010, 45: 519–530
- Iskandarov U, Khozin-Goldberg I, Ofir R, et al. Cloning and characterization of the  $\Delta$ 6 polyunsaturated fatty acid elongase from the green microalga *Parietochloris incisa*. Lipids, 2009, 44: 545–554
- Asamizu E, Nakamura Y, Sato S, et al. A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags. DNA Res, 1999, 6: 369–373
- Asamizu E, Miura K, Kucho K, et al. Generation of expressed sequence tags from low-CO<sub>2</sub> and high-CO<sub>2</sub> adapted cells of *Chlamydomonas reinhardtii*. DNA Res, 2000, 7: 305–307
- Shrager J, Hauser C, Chang C W, et al. *Chlamydomonas reinhardtii* genome project. A guide to the generation and use of the cDNA information. Plant Physiol, 2003, 131: 401–408
- Asamizu E, Nakamura Y, Miura K, et al. Establishment of publicly available cDNA material and information resource of *Chlamydomonas reinhardtii* (Chlorophyta) to facilitate gene function analysis. Phycologia, 2004, 43: 722–726
- Jain M, Shrager J, Harris E H, et al. EST assembly supported by a draft genome sequence: an analysis of the *Chlamydomonas reinhardtii* transcriptome. Nucl Acids Res, 2007, 35: 2074–2083
- Becker B, Feja N, Melkonian M. Analysis of expressed sequence tags (ESTs) from the scaly green flagellate *Scherffelia dubia* Pascher emend. Melkonian et Preisig. Protist, 2001, 152: 139–147
- Simon A, Glöckner G, Felder M, et al. EST analysis of the scaly green flagellate *Mesostigma viride* (Streptophyta): Implications for the evolution of green plants (Viridiplantae). BMC Plant Biol, 2006, 6: 2
- Lanier W, Moustafa A, Bhattacharya D, et al. EST analysis of *Ostreococcus lucimarinus*, the most compact eukaryotic genome, shows an excess of introns in highly expressed genes. PLoS ONE, 2008, 3: e2171
- Shi J, Pan K, Yu J, et al. Analysis of expressed sequence tags from the marine microalga *Nannochloropsis oculata* (Eustigmatophyceae). J Phycol, 2008, 44: 99–102
- Alkayal F, Albion R L, Tillett R L, et al. Expressed sequence tag (EST) profiling in hyper saline shocked *Dunaliella salina* reveals high expression of protein synthetic apparatus components. Plant Sci, 2010, 179: 437–449
- Stanier R Y, Kunisawa M M, Cohen-Bazir G. Purification and properties of unicellular blue-green algae (order Chlorococcales). Bacteriol Rev, 1971, 35: 171–201
- Barnes W M. PCR amplification of up to 35-kb DNA with high fidelity and high yield from  $\lambda$  bacteriophage templates. Proc Natl Acad Sci USA, 1994, 91: 2216–2220
- Wong T K M, Ho C L, Lee W W, et al. Analyses of expressed sequence tags from *Sargassum binderi* (Phaeophyta). J Phycol, 2007, 43: 528–534
- Scala S, Carels N, Galciatore A, et al. Genome properties of the diatom *Phaeodactylum tricornutum*. Plant Physiol, 2002, 129: 993–1002
- Peden J F. Analysis of codon usage. Doctor dissertation. Sutton Bonington: University of Nottingham, 1999
- Tamura K, Dudley J, Nei M, et al. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol, 2007, 24: 1596–1599
- Thompson J D, Gibson T J, Plewniak F, et al. The ClustalX windows interface flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res, 1997, 25: 4876–4882
- Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution, 1985, 39: 66–70
- Campbell C R, Plank C O. Preparation of plant tissue for laboratory analysis. In: Kalra Y P, ed. Handbook of Reference Methods for Plant Analysis. Boca Raton: CRC Press, 1998. 37–49
- Patanjali S R, Parimoo S, Weissman S M. Construction of a uniform-abundance (normalized) cDNA library (reassociation/hybridization/hydroxyapatite). Proc Natl Acad Sci USA, 1991, 88: 1943–1947
- Lambert K N, Williamson V M. cDNA library construction from small amounts of RNA using paramagnetic beads and PCR. Nucleic Acids Res, 1993, 21: 775–776
- Hackett J, Scheetz T, Yoon H, et al. Insights into a dinoflagellate genome through expressed sequence tag analysis. BMC Genomics,

- 2005, 6: 80
- 40 Gurr M I, Harwood J L, Frayn K N. Lipid Biochemistry: An Introduction, 5th ed. Osney Mead, Oxford: Blackwell Science Ltd, a Blackwell Publishing Company, 2002
  - 41 Los D A, Murata N. Structure and expression of fatty acid desaturases. *Biochim Biophys Acta*, 1998, 1394: 3–15
  - 42 Leonard A E, Pereira S L, Sprecher H, et al. Elongation of long-chain fatty acids. *Prog Lipid Res*, 2004, 43: 36–54
  - 43 Yu S Y, Li H, Tong M, et al. Identification of a  $\Delta 6$  fatty acid elongase gene for arachidonic acid biosynthesis localized to the endoplasmic reticulum in the green microalga *Myrmecia incisa* Reisingl. *Gene*, 2012, 493: 219–227
  - 44 Shi X F, Huang J F, Liang C R, et al. Is there a close relationship between synonymous codon bias and codon-anticodon binding strength in human genes? *Chin Sci Bull*, 2001, 46: 1015–1019
  - 45 Grantham R, Gautier C, Gouy M, et al. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res*, 1981, 9: r43–r74
  - 46 Shpaer E G. Constraints on codon context in *Escherichia coli* genes their possible role in modulating the efficiency of translation. *J Mol Biol*, 1986, 188: 555–564
  - 47 Ohno S. Universal rule for coding sequence construction: TA/CG deficiency-TG/CT excess. *Proc Natl Acad Sci USA*, 1988, 85: 9630–9634
  - 48 Andreyeva V M. On the taxonomy of the genus *Myrmecia* Printz (Chlorococcales, Chlorococcaceae) (in Russian). *Bot Mag*, 1978, 63: 442–460
  - 49 Friedl T. Inferring taxonomic positions and testing genus level assignments in coccoid green lichen algae: A phylogenetic analysis of 18S ribosomal RNA sequences from *Dictyochloropsis reticulata* and from members of the genus *Myrmecia* (Chlorophyta, Trebouxiophyceae cl. nov.). *J Phycol*, 1995, 31: 632–639
  - 50 Pröschold T, Leliaert F. Systematics of the green algae: Conflict of classic and modern approaches. In: Brodie J, Lewis J, eds. *Unravelling the Algae: the Past, Present, and Future of Algal Systematics*. Boca Raton: CRC Press, 2007. 123–153
  - 51 Friedl T, Zeltner C. Assessing the relationships of some coccoid green lichen algae and the Microthamniales (Chlorophyta) with 18S ribosomal RNA gene sequence comparisons. *J Phycol*, 1994, 30: 500–506
  - 52 Metting B. The systematics and ecology of soil algae. *Bot Rev*, 1981, 47: 195–312
  - 53 Griffiths D J. The pyrenoid. *Bot Rev*, 1970, 36: 29–58
  - 54 Tschermak-Woess E, Plessl A. Über zweierlei typen der sukzedanen teilung und ein auffallendes teilungsverhalten des chromatophors bei einer neuen Protococcale, *Myrmecia pyriformis* (in German). *Österr Bot Z*, 1948, 95: 194–207
  - 55 Watanabe S, Floyd G L. Comparative ultrastructure of the zoospores of nine species of *Neochloris* (Chlorophyta). *Plant Syst Evol*, 1989, 168: 195–219
  - 56 Dayrat B. Towards integrative taxonomy. *Biol J Linn Soc*, 2005, 85: 407–415

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Supporting Information

**Table S1** NRSs involved in substance and energy metabolism

The supporting information is available online at [csb.scichina.com](http://csb.scichina.com) and [www.springerlink.com](http://www.springerlink.com). The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.